Some problems that were not included on Homework #4

1. (From exam 2, Fall, 2005) An experiment in horticulture investigates the effect of $t = 4$ different herbicide formulations on weed growth in a completely randomized design involving a total of 20 crabgrass plants. Let y_{ij} denote the observed shoot dry weight (in cg) of the j^{th} plant 120 days after application of treatment i. Summary statistics below:

Treatment i	Herbicide Formulation	sample size	sample mean, $\bar{y}_{i+}$	sample std. dev
1	Control	5	300	80
2	A300	5	260	70
3	A400	5	240	60
4	Surflan	5	200	70

(a) Consider another parameterization for the (20×1) vector Y: $Y = X\beta + E$ where

$$X = \begin{pmatrix} 1 & 2 & 0 & -1 \\ 1 & 2 & 0 & -1 \\ 1 & 2 & 0 & -1 \\ 1 & 2 & 0 & -1 \\ 1 & 2 & 0 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & 0 & 0 & 3 \\ 1 & 0 & 0 & 3 \\ 1 & 0 & 0 & 3 \\ 1 & 0 & 0 & 3 \\ 1 & 0 & 0 & 3 \end{pmatrix}$$

i. Let y_{i+} denote the i^{th} treatment *sum*. Express the matrix product $X'Y$ in terms of $y_{1+}, y_{2+}, y_{3+}, y_{4+}$.

ii. Let $\bar{y}_{i+}$ denote the i^{th} treatment *mean*. Obtain the least squares estimate, $\hat{\beta}$, of the vector of regression coefficients, in terms of $\bar{y}_{1+}, \bar{y}_{2+}, \bar{y}_{3+}, \bar{y}_{4+}$.

iii. Obtain $X'X$.

iv. Obtain the variance-covariance matrix for $\hat{\beta}$

v. In terms of shoot weights, what is being estimated by $\hat{\beta}_2$?

vi. In terms of shoot weights, what is being estimated by $\hat{\beta}_3$? (Note that $\beta_1, \beta_2, \beta_3$ are $4-1=3$ orthogonal contrasts, that is, contrasts so that $X'X$ is a diagonal matrix and the three estimates $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ are uncorrelated. Note that the regression sum of squares is equal to the treatment sum of squares. This is true of any design matrix X of full rank.

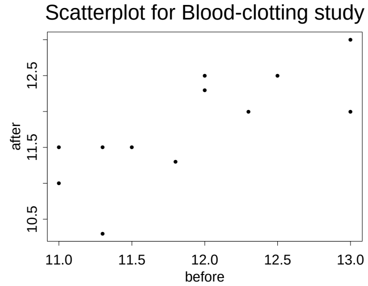
vii. What is the correlation between $\hat{\beta}_2$ and $\hat{\beta}_3$?

(b) Consider the model $Y_{ij} = \mu + \tau_i + E_{ij}$ where E_{ij} are iid $N(0, \sigma^2)$. Derive expressions for $E(MS[Trt])$ and $E(MS[E])$ under this model.

2. A person's blood-clotting ability is typcially expressed in terms of a "prothrombin time," which is defined to be the interval between the initiation of the prothrombin-thrombin (two proteins) reaction and the formation of the final clot. Does *aspirin* affect this function? Measurements made before administration of two tablets and three hours after.

	Prothrombin Times (seconds)		
Subject	Before Aspirin (Y_1)	After Aspirin (Y_2)	Difference (D)
1	12.3	12.0	0.3
2	12.0	12.3	-0.3
3	12.0	12.5	-0.5
4	13.0	12.0	1.0
5	13.0	13.0	0.0
6	12.5	12.5	0.0
7	11.3	10.3	1.0
8	11.8	11.3	0.5
9	11.5	11.5	0
10	11.0	11.5	-0.5
11	11.0	11.0	0
12	11.3	11.5	-0.2

Scatterplot for Blood-clotting study



(a) Carry out a paired t-test of the hypothesis that prothrombin time is unaffected by aspirin.

(b) Carry out an F-test of the same hypothesis treating subjects as blocks in an analysis for a RCBD.

(c) Show that, in general, the paired t-test is equivalent to the F-test for the RCBD with block size equal to 2.

3. (taken from Ott and Longnecker 15.10, p. 889) Fuel efficiency of four blends of gasoline is measured in MPG. There is considerable variability due to driver Another source of variability is model of car. An experiment randomizes four models of car and gasoline blends (A,B,C,D) to drivers according to the design below

	Model			
Driver	1	2	3	4
1	15.5(A)	33.8(B)	13.7(C)	29.2(D)
2	16.3(B)	26.4(C)	19.1(D)	22.5(A)
3	10.5(C)	31.5(D)	17.5(A)	30.1(B)
4	14.0(D)	34.5(A)	19.7(B)	21.6(C)

(a) Assuming normally distributed data, propose a model in which the effects of model, driver and blend are additive on the mean.

(b) Report estimates for these parameters.

(c) Using $\alpha = 0.05$, test the effects of each experimental factor.

4. (This problem is fun.) Click the link on the ST511 website that takes you to the page of applets by Webster at Univ. of South Carolina. Then click the "Let's Make a Deal" link.

`www.stat.tamu.edu/~west/applets/LetsMakeaDeal.html`.

Read the description of the problem. Now, consider an experiment where you get to observe n independent $0-1$ trials and you are interested in competing hypotheses for the success probability p:

$$H_0 : p = 1/2 \quad \text{versus} \quad H_1 : p = 2/3$$

Let Y denote the *number* of successes out of the n trials.

(a) Suppose you observe $n = 30$ such trials and adopt this critical region for Y :

$$\text{Reject } H_0 \text{ if } Y \geq 20.$$

Using Table C.5 or an appropriate BINOMIAL applet, obtain the exact significance level of this test.

(b) Using the applet by Lenth, find either the approximate or exact power of the test which uses the critical region we've specified.

`www.stat.uiowa.edu/~rlenth/Power/`.

(c) Collect your data using the switch strategy. Play the game $n = 30$ times and report your results. Are your data "statistically significant", in terms of the test you've set up? Briefly state your conclusion regarding H_0 and H_1, using α from part a).

5. Visit the website

`http://www.ruf.rice.edu/~lane/stat_sim/two_way/index.html`

read the instructions and give answers to questions 1-8. Wherever you're asked to "make up data", simply provide treatment means. In questions 6-8, there is disagreement between the applet and the exercises. $A \times B$ are arranged 3×2 in the former, but 2×3 in the latter. So, switch "A" and "B" in 6-8.

6. Consider the 21-day chicken bodyweights from class notes on p.116.

   (a) Carry out all pairwise comparisons among the four treatment means. Use Tukey's procedure with $fwe = 0.05$. Provide an interpretation of fwe. Obtain simultaneous 95% confidence intervals for all pairwise differences.

   (b) "omitted exam questions from p.124 of lecture notes": problems 1,2abcde,3ab.

   (c) Remark on the appropriateness of the computations in problem (a) in light of the results of the lack-of-fit test for the linear regression model.

7. Consider a completely randomized design for an experiment to test the effectiveness of three drugs (A,B,C) in dissolving blood clots. Suppose that the mean times required to stop the bleeding in a cut of specified size are $\mu_A = 150, \mu_B = 140, \mu_C = 160$.

   (a) Specify the sampling distribution of $F = MS(Trt)/MS(E)$ when the standard deviation of times for a given drug are $\sigma = 10$. Do the same when $\sigma = 20$.

   (b) Obtain the sample size necessary achieve a power of 0.9 to declare the difference among drug sample means to be significant in the case where $\sigma = 10$. Do the same when $\sigma = 20$.